A Simulation Approach to Modelling Temporary Populations

Elin Charles-Edwards, University of Queensland e.charles-edwards@uq.edu.au ;
Martin Bell, University of Queensland, martin.bell@uq.edu.au

Paper prepared for the XXVI IUSSP Conference, Marrakech, Morocco, 30th September – 2nd October 2009

Abstract

Conventional population estimates provided by statistical agencies for regions and localities refer to a single point in time but it is widely recognised that actual population numbers fluctuate over the course of a year. Direct measurement of temporary populations has proven prohibitively expensive, while indirect estimates based on symptomatic data have met with limited success. We propose an alternative approach based on a stochastic simulation model that couples seasonality of mobility with duration of stay and is made operational using Monte Carlo methods. Drawing on data from the Australian National Visitor Survey, we identify families of distributions fitting seasonality and duration. Random samples are then drawn from these, and combined to produce estimates of 'Visitor Nights' for thirteen selected municipalities. Results are promising with the synthetic estimates correlating closely with observed data from the Australian Small Area Accommodation Survey.

1. Introduction

Official population estimates for Australia, as in many parts of the world, are premised on the demographic accounting equation which measures population growth (or decline) between periods based on the births, deaths and migration occurring within the estimation interval. Absent from these estimates is any indication of the short-term shifts in population driven by temporary population mobility – those geographical moves more than one night in duration that do not entail a change in usual residence (Bell and Ward 1998).

The scale of temporary population mobility in Australia is extraordinary. On the night of the 2006 Australian Census of Population and Housing, almost one million Australians, one person in 20, were away from their place of usual residence. This is a fraction of the estimated 285 million nights spent away from home by Australians aged 15 and over in 2006 (NVS 2007). At the local level, this mobility translates into significant fluctuations in population numbers over the course of the year. The size of temporary inflows, and the demand for goods and services generated by visitor populations, has lead to calls for population estimates incorporating these non-resident populations (Rose and Kingma 1989; ABS 1996).

Researchers interested in the estimation of non-resident populations have invested significant energy in the search for data capturing, either directly or indirectly, the temporary populations of regions (Smith 1989; Lee 1999). The consensus is that extant data on temporary population mobility is sparse, fragmented, and generally deficient for the purpose of population estimation. There is an associated sense that, in the absence of major improvements in data quality, reliable estimates of temporary populations will remain unattainable.

A contrary view is that data paucity is not necessarily an obstacle to the estimation of temporary populations. Migration researchers have long made use of instruments such as migration age schedules to infer patterns of permanent migration in the absence of comprehensive data (Castro and Rogers 1983), and multiregional models of migration are frequently simplified via the parameterization, aggregation and partitioning of data (Rees 1997; Wilson and Bell 2004). Other subfields of demography respond to incomplete and missing data by using instruments such as model fertility schedules and life tables (Judson and Popoff 2004).

Compared with other components of demographic change, our understanding of temporary population mobility remains rudimentary. Nevertheless, progress has been made, both in understanding the scale and composition of temporary mobility (Bell and Ward 2000; Bell 2004; Brown and Bell 2005), and in exploring its temporal and spatial dimensions (Bell 2004). Drawing on this work, we propose an approach to the estimation of temporary populations in individual destination regions based on modelling key elements of the Australian temporary mobility system. The model is estimated as a stochastic simulation using data from the Australian National Visitor Survey.

The paper begins by defining temporary population mobility, before outlining the process whereby an initial conceptualisation of the Australian temporary mobility system is transformed into a model estimating the temporal variability in visitor populations within individual destination regions. The sequence of model reductions seeks to preserve the architecture of the Australian mobility system, whilst increasing parsimony, so that the final model can be estimated using data from the Australian National Visitor Survey and Australian Survey of Small Area Accommodation.

Following the discussion of the modelling approach and data employed in this study, the analysis shifts to parameterisation of model terms representing the seasonality and duration of temporary moves to destination regions. The model is calibrated for Australian Tourism Regions over the period 2002-06 before estimates are generated for thirteen of Australian Local Government Areas. The paper concludes with a summary of key findings and a discussion of the potential for similar modelling approaches to the estimation of temporary populations in other geographic settings.

2. Modelling temporary population mobility

For this paper, temporary population mobility is defined as a geographical move involving a stay of one night or more away from the place of usual residence, but less than one year in duration. As is the case for permanent migration, geographical thresholds of temporary movements are dictated by the available data (de Gans 1994). Here, these limits extend from a minimum of 40 kilometres to a maximum distance of around 6000 kilometres, defined by the scope of the National Visitor Survey (NVS) and the physical extent of the Australian continent. Temporary moves can be undertaken for a variety of purposes which manifest in a multitude of space-time behaviours, all with the aim of meeting *goals in space and time* (Roseman 1992; Hooimeijer and van der Knaap 1994). Examples range from brief visits to friends or family, to the cyclical rosters of Fly-in Fly-out miners (Houghton 1993), and from long-distance travel to access health care, to the annual family holiday at the beach.

The cumulative outcome of temporary mobility at the system level is an undulating population surface in which, waves, currents and tides of movements produce a continuous ebb and flow of people across the Australian continent. Underpinning this population surface are sets of spatial interactions between origins and destinations, which shift according to characteristics at the origin and destination, which likewise vary over time. This population surface can be further envisaged as stratified according to purpose, with absolute and relative shifts in the size of the layers producing the undulations visible at the surface. The size of peaks can vary from minor peaks in visitor numbers to periodic surges threatening to overwhelm host communities (Charles-Edwards, Bell, and Brown 2008).

Transformation of this description of the temporary population system into a model estimating visitor numbers within individual destination regions represents a significant reduction in information from the system. A key decision is whether to model the system at the micro- or macro-level. Micro-level studies of permanent migration have tended to focus on the decision of individuals to migrate, whilst macro-level studies have sought to answer question relating to the regional and national drivers of aggregate migration flows (Stillwell and Congdon 1991). In recent times attempts have been made to bridge this macro-micro divide using micro-simulation and agent-based models integrated with macro-level spatial models of migration processes (Wu, Birkin, Rees 2008) but the data demands of such models are high. Both micro and macro approaches to the study of spatial mobility seek to represent the same underlying processes (Van Imhoff and Post 1998), and the choice of approach is ultimately driven by model purpose. In the present study a macro-level approach to the mobility system has been adopted for two reasons: (i) the lack of empirical knowledge of the temporary mobility system; (ii) the need to produce destination based models of temporary populations based on regional characteristics.

A simplified macro-level representation of a temporary mobility system can be achieved by the addition of a time dimension to a spatial interaction model, reflecting the earlier conceptualisation of temporary population mobility as a means of meeting *goals in space and time* (Equation 1). In this formulation the intensity of temporary flows between origins and destinations reflects their relative attractiveness, which can shift over time. Estimates of visitor populations can then be derived by calculating the nightly stock of temporary visitors at individual destination regions.

$$ODT$$

Where: O is the origin

D is the destination

T is time

[Equation 1]

While this formulation encompasses the complete detail of the temporary mobility system, the data requirements for estimating the fully saturated model are high. To estimate movers across the 84 origin-destination pairs, which comprise the principle geography used in this study, at nightly intervals, to reflect the minimum duration of temporary moves, requires around 2.6 million variables. Data at this temporal and spatial resolution are not currently available in Australia, and are not likely to become so in the foreseeable future.

Excessive data demands are a persistent issue in multi-regional migration modelling (van Imhoff, van der Gaag et al. 1997). Rees (1997) suggested that the multiregional migration model, and by extension the temporary model, could be shrunk in three ways:

1. Aggregation, in which the number of variable categories is reduced (e.g. compression of single year age groups to five year age groups)
2. Partitioning, whereby the fully saturated model is separated into a number of sub-models (e.g. Migrant Pool Model(Bell and Wilson 2004)); and
3. Parameterization, whereby observed data are replaced by a smaller number of parameters (e.g. Model Migration Schedules (Rogers, Raquillet et al. 1978)).

Such restrictions increase model parsimony while leaving the essential architecture of the model unchanged (van Imhoff, van der Gaag et al. 1997) but there inevitably are sacrifices in terms of reduced model fit. Aggregation of the model specified in Equation 1 can be achieved in two ways: by combining the spatial units, thus reducing the number of interactions being modelled; or temporally, by increasing the interval over which intensity is measured. Since the aim here is to produce estimates of visitor populations within individual destination regions, there is little scope to reduce the spatial scale of the model.

It is, however, possible to make adjustments in respect to the time dimension by increasing the interval at which the model is run.

The chosen interval must reflect the temporality of temporary population movements. Many forms of temporary population mobility in both the developing (Chapman and Prothero 1983); and developed world contexts (Hanson and Bell 2007) are characterised by seasonal cycles, measured by reference to monthly intervals. Examples include the mobility of elderly snowbirds (Hogan and Steinnes 1993); second home owners, harvest labourers (Hanson and Bell 2007) and tourism (BarOn 1975; Lundtorpe 2001; Koenig-Lewis and Bischoff 2005). Weekly cycles are also observed in a number of forms of temporary mobility (e.g. Fly-in Fly-out miners), but the former is selected for this model because, in spite of the increasingly routine nature of some forms of temporary mobility, the majority of temporary moves are posited to occur outside the weekly routine of work and leisure(Hall 2008) .

A consequence of adjusting the model time interval is the need to shift from modelling temporary movers to moves. In the analysis of permanent migration it has been found that as the length of interval over which migrations are measured increases, the gap between the number of movers and the number of moves widens (Bell, Blake et al. 2002). Thus, if the full impact of temporary mobility at destination regions is to be captured, the model needs to be estimated based on the number of moves, as opposed to the number of movers. A second consequence of adjusting the time interval is a requirement to include a measure of movement duration. Because temporary moves are of variable duration, the number of moves occurring in an interval does not reflect the cumulative or effective visitor population in a destination region. Rather, visitor populations vary proportionally with both the number of visits to a destination and with the average duration of stay. That is to say, a doubling of either the number of visits or the average duration of visit to a destination would result in a doubling of the effective visitor population over a given interval.

In Equation 2 the basic spatial interaction architecture of Equation 1 is preserved, but with a substitution of the time variable by a seasonality term and a duration term. The seasonality term (S) represents the monthly intensity of temporary mobility in the system, and duration (D), the cumulative duration of moves over monthly intervals. The output from this model is an estimate of Visitor Nights, by month, across the mobility system. This substitution significantly reduces the data demands for this model from an estimated 2.6 to around 1 million variables for 84 tourism regions; however, the data demands from this model are still too high for practical implementation in the Australian context.

$$OD\delta S$$

Where:

O	is the origin
D	is the destination
S	is the seasonality of temporary moves (monthly intensity)
δ	is duration

[Equation 2]

An avenue for further reduction of the model can be made in light of its final function, that is, to estimate the temporal variability in visitor numbers to individual destination regions. Just as it is not always necessary (or desirable) to run full multi-regional migration models to estimates inflows to individual destinations, it can be argued that a fully saturated model is not necessary (or desirable) in the production of estimates of temporary populations within individual destination regions. This is particularly true if the model is to have wider application by non-expert users, such as local planning officials, who require estimates of visitor populations within for their local area. One way in which the model may be usefully reduced consistent with this imperative is by partitioning it into origin and destination terms such that:

$$OD + (sO + \delta O) + (sD + \delta D)$$

Where:

O	is the origin
D	is the destination
δ	is duration
s	is the seasonality of temporary moves (monthly intensity)

[Equation 3]

In this formulation, origin and destination interactions are restricted to aggregate moves between regions. Both seasonality and duration are allowed to vary independently of one another at the origin. Similarly, seasonality and duration vary independently with the destination, but again there is no interaction between these terms. Restriction of the model in this way leads to a 92: 1 reduction in the number of variables compared with the fully saturated model. The key benefit arising from this formulation is that the overall intensity of mobility (OD) can be modelled independently of the temporal patterning of this mobility at the origin and destination ($(sO+ \delta O) + (sD +\delta D)$). This is critical, as temporal variability remains a major barrier in the estimation of temporary populations, and a key motivation underpinning indirect approaches to the estimation of temporary populations via symptomatic data such as electricity usage (Smith 1989).

A final model reduction is the exclusion of the origin seasonality and origin duration terms. Whilst this represents a significant loss of information from the model, earlier model partitioning makes these terms redundant for the purpose of estimating visitor populations in individual destination regions. This leaves Equation 4 below:

$$OD + sD + \delta D$$

Where:	O	is the origin
	D	is the destination
	S	is seasonality
	δ	is duration

[Equation 4]

This final model formulation and represents a 284: 1 reduction in variables when compared with the fully saturated model specified in Equation 1. Ideally, estimates of the degree of information lost in these reductions would be made, however, data quality severely limit the potential for this in the Australian case.

The task now is to translate this conceptual model into an operational form that can be estimated using available data.

3. Data

Compared with other industrialised countries, Australia is well served by data on temporary population mobility. There are three large-scale data sets collecting information on temporary movements: the Australian Census of Population and Housing; Australian Survey of Tourism Accommodation; and the National Visitor Survey). While each data set has some analytical value, the NVS is by far the most comprehensive when measured in terms of capturing temporary population mobility over time and is the only source that collects data across the four model elements (Table 1).

Table 1: Data sources on temporary population mobility, Australia

	Intensity	Origin (O)	Destination (D)	Duration (δ)	Seasonality (S)
Census	✓	✓	✓		
STA	✓		✓	✓	✓
NVS	✓	✓	✓	✓	✓

The principal Australian data source collecting information on temporary population mobility in Australia is the Australian National Visitor Survey (NVS). Operating in its current guise since 1998, the survey currently samples around 120,000 Australians aged 15 years and over, every year, on their domestic travel behaviour over the previous four weeks. The sample size has increased from 80,000 from in January 2005 to improve sample estimates across small regions. Respondents are questioned on their domestic overnight travel with information procured on the location, timing, and purpose of travel, is collected along with data on expenditure, accommodation, transport and other trip characteristics. Data are weighted to provide estimates of total travel behaviour based on the age, sex, household size, region and month of travel along with the length of visits and benchmarked against the Australian population age 15 years and over (TRA 2009).

Despite its utility, the NVS has several limitations. Being a dwelling based survey, more mobile sectors of the population, such as ‘Grey Nomads’ who spend much of the year travelling around Australia (Mings 1997), are likely to be under enumerated. There are also issues with the enumeration of persons with what can be classed as multiple dwellings , for example Fly-in Fly-out miners who live in employee supplied housing. The key issue with these data, however, is the sampling variability within these data, a particularly acute issue in this study with a focus on visits to regions over short time intervals. Table 2 shows the confidence intervals for NVS data at the 95% predictive interval. Two sets of confidence intervals are presented, owing to the increase in sample size that occurred in January 2005 from 80,000 to 120,000 Australians aged 15 and over.

Table 2: Confidence Intervals for NVS Estimates at the 95% level

	1998-2004		2005 onwards	
Estimates (000s)	Overnight visitors	Visitor Nights	Overnight visitors	Visitor Nights
20	>50	>100	>50	>100
50	>50	>100	47.5	>100
80	41.5	>100	37.7	>100
100	37.5	>50	33.8	>100
200	27.4	>50	24.0	>50
300	22.8	>50	19.6	>50
500	18.1	48.3	15.3	45.7
1000	13.2	35.8	10.8	32.4
2000	9.7	26.6	7.7	23.0
3000	8.1	22.3	6.3	18.8
5000	6.4	17.9	4.9	14.6
7000	5.5	15.5	4.2	12.3
10000	4.7	13.3	3.5	10.3
20000	3.4	9.8	2.5	7.3
30000	2.8	8.3	2.0	6.0
50000	2.3	6.6	1.6	4.6
70000	1.9	5.7	1.3	3.9
100000		4.9	1.1	3.3
140000		3.6	0.9	2.8
200000			0.8	2.3
500000			0.5	1.5
1000000				1.0

(NVS 2004, 2005)

Some indication of the severity of sampling variability can be gleaned by examining the range of estimates across regions in a single year. Looking at total trips undertaken to 84 Australian Tourism Regions in 2006, 70.2% record values less than 1000 at which the 95% confidence interval records a value of 10.8%, moreover, when annual figures are disaggregated by month, values in more than 96% of regions fall below this threshold. Estimates for Visitor Nights (the metric in which final estimates of temporary populations are produced) are likewise characterised by a high degree of sampling variability at the regional level, with 96% of regions recording annual estimates with confidence intervals greater than 10.3%. The sampling variability in NVS data preclude its use in the direct estimation of visitor populations; however, the data do allow the model terms specified in Equation 4, to be estimated for most of the 84 Australian Tourism regions, albeit with some aggregation.

A second source of data measuring a subset of temporary moves is the Australian Survey of Tourism Accommodation (STA) administered by the Australian Bureau of Statistics. These data provide information on the supply of, and demand for temporary accommodation in '…hotels, motels and guest houses and serviced apartments with 5 or more rooms or units; holiday flats, units and houses of letting entities with 15 or more rooms or units; caravan parks with 40 or more powered sites and visitor hostels with 25 or more bed spaces' (Australian Bureau of Statistics 2005). The survey is conducted as a census of approximately 4000 short-term accommodation units, limited to:

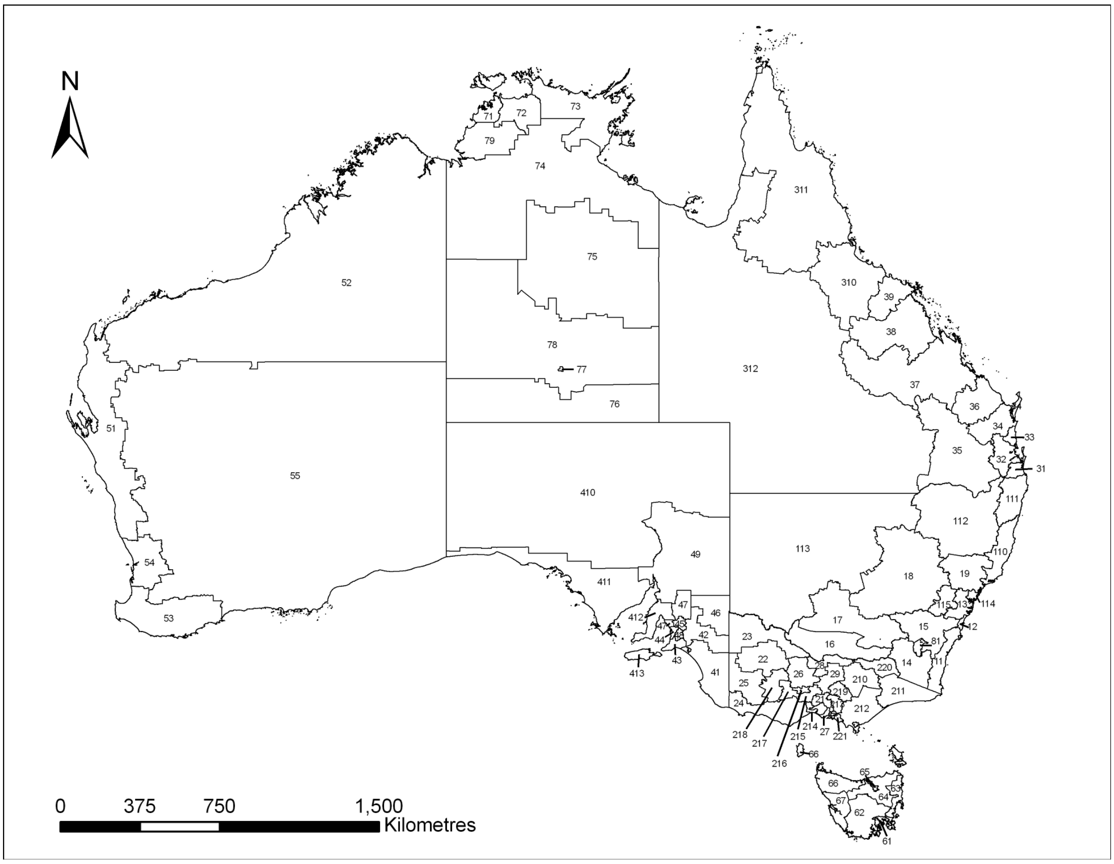
- *licensed hotels and resorts with facilities and 5 or more rooms*
- *motels, private hotels and guest houses with facilities and 5 or more rooms*
- *serviced apartments with 5 or more units*
- *caravan parks with 40 or more powered sites*
- *holiday flats, units and houses of letting entities with 15 or more rooms or units*
- *visitor hostels with 25 or more bed spaces.*

(ABS 2009)

Variables collected include the number of monthly arrivals and length of stay as well as items on visitor expenditure. While these data provide some useful information on the space-time dynamics of temporary mobility, their restricted sampling frame limits their utility at the local level. The can, however, when adjusted via Iterative Proportional Fitting using NVS data, assist in the evaluation of model outputs for Local Government Areas.

4. Geography

Figure 1 shows the principal geography, Tourism Regions, used in this study.



11	South Coast	29	Goulburn	311	Tropical North Queensland	64	Northern
12	Illawarra	210	High Country	312	Outback	65	Greater Launceston
13	Sydney	211	Lakes	41	Limestone Coast	66	North West
14	Snowy Mountains	212	Gippsland	42	Murraylands	67	West Coast
15	Capital Country	213	Melbourne East	43	Fleurieu Peninsula	71	Darwin
16	The Murray	214	Geelong	44	Adelaide	72	Kakadu
17	Riverina	215	Macedon	45	Barossa	73	Arnhem
18	Explorer Country	216	Spa Country	46	Riverland	74	Katherine
19	Hunter	217	Ballarat	47	Clare Valley	75	Tablelands
110	North Coast NSW	218	Central Highlands	48	Adelaide Hills	76	Petermann
111	Northern Rivers - Tropical NSW	219	Upper Yarra	49	Flinders Ranges	77	Alice Springs
112	New England North West	220	Murray East	410	Outback SA	78	MacDonnell
113	Outback NSW	221	Phillip Island	411	Eyre Peninsula	79	Daly
114	Central Coast	31	Gold Coast	412	Yorke Peninsula	81	Canberra
115	Blue Mountains	32	Brisbane	413	Kangaroo Island		
21	Melbourne	33	Sunshine Coast	51	Australia's Coral Coast		
22	Wimmera	34	Hervey Bay/Maryborough	52	Australia's North West		
23	Mallee	35	Darling Downs	53	Australia's South West		
24	Western	36	Bundaberg	54	Experience Perth		
25	Western Grampians	37	Fitzroy	55	Australia's Golden Outback		
26	Bendigo Loddon	38	Mackay	61	Greater Hobart		
27	Peninsula	39	Whitsundays	62	Southern		
28	Central Murray	310	Northern	63	East Coast		

Figure 1: Australian Tourism Regions 2005

Tourism Regions, used in the dissemination of NVS and STA data, are a specialised geography based on an aggregation of Australian Statistical Local Areas which in turn are part of the main structure of the Australian Standard Geographical Classification. Concordance files between Australian Tourism Regions and Statistical Local Areas are published annually by the Australian Bureau of Statistics. As of June 2008, there were 85

Tourism Regions. As much of the analysis of the spatial and temporal structuring of temporary population mobility across Australia is conducted for the period 2002-06, the 2005 Tourism Region geography is used. Data for 1998-2006 are referenced to this geography.

5. A model estimating Visitor Nights in Australian regions

The model specified in Equation 4 was comprised of three terms: the first (OD) reflecting the moves between origin and destination pairs; the second (SD), the seasonality of visitation to regions; and a third (δD), the duration of stay at the destination regions. The model is formalised as a stochastic model, with the output measured in Visitor Nights to destination regions due to the inclusion of a duration term in the model. Visitor Nights can be readily converted into a full-time equivalent population measure by dividing it by the number of nights within a given interval, thus making it comparable to estimates of resident populations.

Equation 5 sets out the formal model for estimating Visitor Nights at the destination.

$$VN_t(a,b) = \sum_{i=1}^{i=M*m_t(a,b)} d_t(a,b)$$

Where VN (a, b) is the probability distribution for Visitor Nights in a destination region

t is the month

m (a,b) is the probability distribution reflecting the seasonality of visits to a region

d (a, b) is the probability distribution for the duration of stay at a destination

[Equation 5]

The model is specified as a stochastic Monte Carlo simulation to incorporate the uncertainty associated with the underlying dimensions of mobility and propagate these to the final estimates. The Monte Carlo method simulates statistical processes by the empirical action of repeatedly drawing random samples from known distributions (Mooney 1997), and is chosen over here traditional methods of statistical inference because of the flexibility this method provides in dealing with non-traditional input distributions[1].

The model procedure has two stages: first, estimates of the number of visits to a destination region, in given month, are generated from an input distribution reflecting the monthly share of annual visitation to said region. This determines the size of the sample to be drawn from a corresponding probability distribution of visit durations for that region, in a given month. Sampled duration values are then summed to produce the random variate, Visitor Nights. This process is repeated until mean Visitor Nights converges, measured as a less than 1% variation in the sample statistic. The model output is a sampling distribution of Visitor Nights for an individual region, for a given month. A complete annual profile of Visitor Nights is generated by running the model for all months.

These simulations are programmed in Microsoft Excel ™ 2007 using Visual Basic for Applications (VBA). This is a popular platform for simulation (Keeling 2004) due to its transparency and relative simplicity for end users of system. Pseudo-random numbers are generated in VBA using the 'Rnd' command which samples variates from the unit uniform distribution using a deterministic algorithm. There are issues with the randomness of Excel's pseudo-random number generators (McCullough 2008) which preclude their use in some types of simulations; however, with a period of approximately 16.5 million numbers it is adequate for our modelling purposes.

The success of the model specified in Equation 5, is contingent on the identification of appropriate sampling distributions for seasonality and duration at individual destination regions. This was done as part of a wide-ranging analysis of the spatial and temporal patterns of Australian temporary population mobility and the determinants underpinning this system, not elaborated here[2].

---

[1] The full range of input distribution experimented with in the simulation model are not elaborated here

[2] Results of these analyses are forthcoming

Distributions for seasonality were specified based on observed spatial regularities in seasonality across Australian regions. A *k* means cluster analysis, a widely used classificatory technique, was undertaken using z-score standardized Seasonal Factors for tourism regions as the clustering criteria. These Seasonal Factors, a measure of the seasonal component of time series, were generated via Census X-11 time series decomposition of NVS data for the period 2002 through 2006.The outcome of these analysis were four classes of regions with distinct seasonal profiles of visitation.

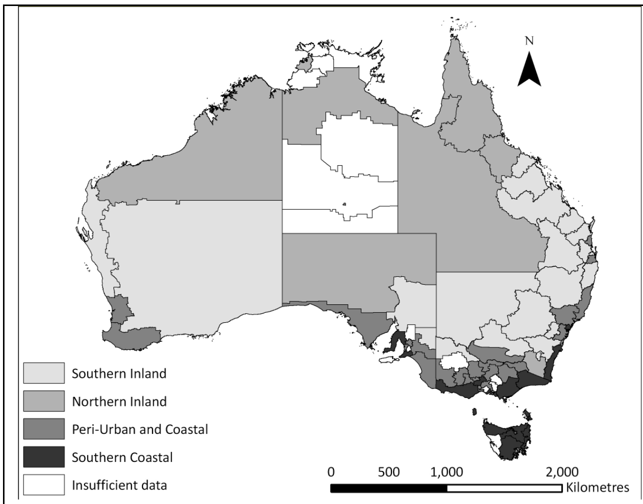


Figure 2: Regional Groupings, Seasonality

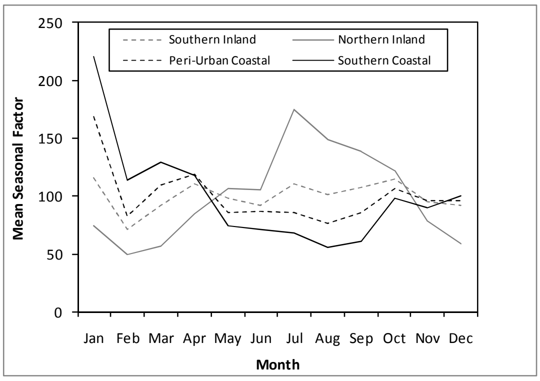


Figure 3: Regional Groupings, Seasonal Profiles

Marked spatial differentiation exists between the four seasonality groupings. The *Southern Inland* and *Peri-Urban and Coastal* groupings are the least seasonal of the four groups, with visitation to both sets of regions characterised by relatively modest peaks in January, April and October, mirroring the time of Australian school holidays. The January peak in visitation is higher in *Peri-Urban and Coastal* regions reflecting the preference for the beach as an Australian summer holiday destination, whilst *Southern Inland* Regions experience higher levels of visitation during the winter months, reflecting the more amenable climatic conditions at this time. *Southern Coastal* and *Northern Inland* regions demonstrate even higher levels of seasonality, with visitation to *Southern Co*astal regions concentrated in the summer months, falling off drastically in the winter months, whilst the inverse is true for *Northern Inland* regions. Again these profiles reflect the relative climatic attractiveness of regions at these times, but also the composition of movements to these regions.

The seasonal profiles of each of these four regional grouping was translated into a set of twelve normal sampling distributions estimated from the Seasonal Factors of individual member regions. The mean and standard deviations of these distributions are shown in Table 3.

Table 3: Normal distribution parameters, Seasonality

		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Southern Inland	μ	116.0	70.9	91.6	110.7	97.9	91.6	110.6	101.3	107.3	115.0	94.6	91.7
	σ	14.1	12.2	19.8	16.0	12.9	13.1	17.2	21.1	21.4	21.9	15.8	15.5
Northern Inland	μ	74.1	49.1	57.1	85.1	106.4	105.8	174.3	149.0	138.4	121.5	78.2	59.3
	σ	19.9	13.3	16.2	21.7	29.2	28.8	23.6	30.1	14.6	32.1	11.9	16.1
Peri-urban and Cosatl	μ	168.3	82.2	109.9	119.2	86.3	86.9	85.3	76.5	86.0	106.0	96.6	96.1
	σ	19.4	13.3	22.0	18.1	13.7	11.1	14.0	10.6	12.4	17.7	14.1	15.4
Southern Coastal	μ	220.8	113.5	128.8	117.8	74.1	71.5	67.9	55.8	60.7	97.8	89.9	100.7
	σ	28.0	25.4	32.7	6.0	10.3	11.8	13.6	9.4	5.0	21.3	11.5	13.9

Families of distributions of visit durations to regions were indentified in a similar fashion, using a *k* means cluster analysis of standardised mean visit durations (Figures 4 and 5). Five classes of regions were identified underpinned by a spatial structure reflecting the relative accessibility of regions. Seasonality is present in all duration profiles, yet differs in both degree and pattern across clusters.

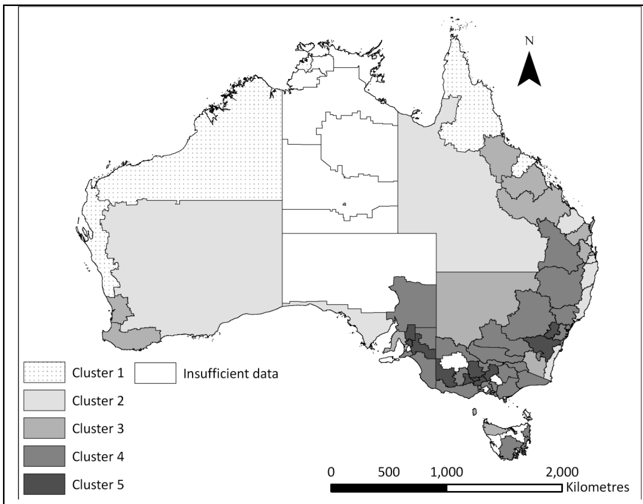


Figure 4: Regional Groupings, Duration

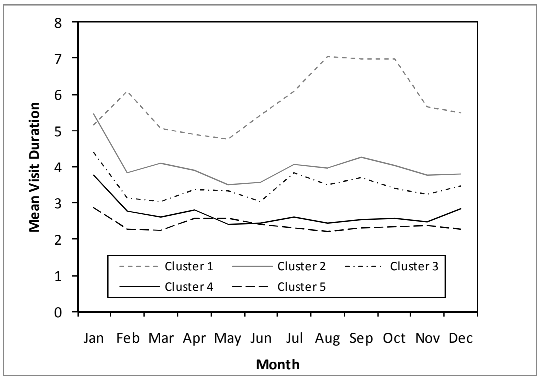


Figure 5: Regional Groupings, Mean Duration Profiles

Sampling distributions were parameterised for these regional groupings by month using the proprietary distribution fitting software, Easyfit ™. A number of distribution types were experimented with in the final simulation including the Weibull distribution and a multinomial distribution, but only results from the lognormal distribution, the best performing distribution overall are presented here. The parameters of these sampling distributions are shown in Table 4.

Table 4: Lognormal distribution parameters, Duration

		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Cluster 1	σ	0.89	0.98	0.89	0.81	0.89	0.89	0.92	0.97	0.97	0.95	0.94	0.85
	μ	1.26	1.15	1.10	1.24	1.15	1.22	1.38	1.44	1.42	1.44	1.23	1.17
Cluster 2	σ	0.91	0.81	0.79	0.76	0.77	0.74	0.81	0.86	0.86	0.79	0.79	0.79
	μ	1.31	0.95	0.97	1.05	0.89	0.91	1.01	0.96	1.02	1.08	0.96	1.02
Cluster 3	σ	1.13	0.77	0.73	0.76	0.79	0.72	0.79	0.77	0.82	0.78	0.77	0.77
	μ	0.90	0.80	0.83	0.88	0.83	0.81	0.91	0.86	0.90	0.90	0.83	0.87
Cluster 4	σ	0.83	0.70	0.66	0.69	0.67	0.64	0.70	0.67	0.68	0.69	0.67	0.72
	μ	0.96	0.70	0.69	0.77	0.65	0.67	0.71	0.65	0.69	0.72	0.67	0.77
Cluster 5	σ	0.72	0.62	0.62	0.66	0.63	0.62	0.63	0.61	0.63	0.65	0.64	0.63
	μ	0.70	0.54	0.57	0.68	0.60	0.61	0.58	0.55	0.57	0.64	0.59	0.59

The distributions specified in Tables 3 and 4 are the basic input into the Monte Carlo simulation of visitor populations to regions. Distributions are assigned to regions based on cluster membership.

The model was initially calibrated for Tourism Regions for the period 2002-2006, for 66 regions, with modelled distributions of Visitor Nights, evaluated against observed values of monthly Visitor Nights to regions as measured by the NVS. In the initial model run, 87.4% of observed monthly values fell within the 95% predictive interval generated by the model. Due to the sampling variability in the observed data, this is conservative estimate of model fit. When the sampling variability of the observed data was incorporated into the model evaluation, observed values for only two regions fell outside the 95% predictive interval. Figures 6 and 7 show model results for two regions, South Coast (NSW), and Northern (Qld), in illustrate the fit of the modelled profiles. Inspection of these graphs suggests a good fit between the modelled and observed data, and critically a strong concordance between the shapes of the modelled and observed profiles.

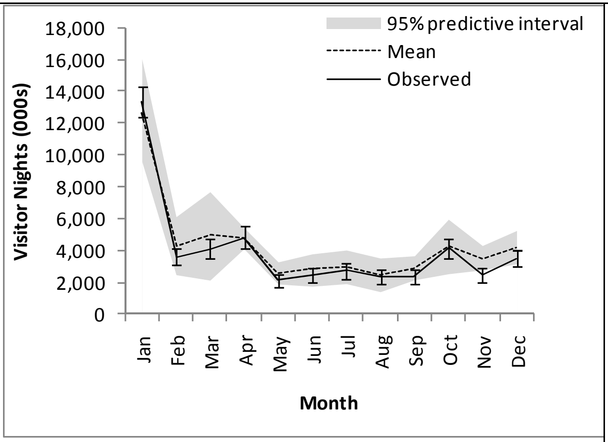


Figure 6: Modelled versus observed data, South Coast (NSW)

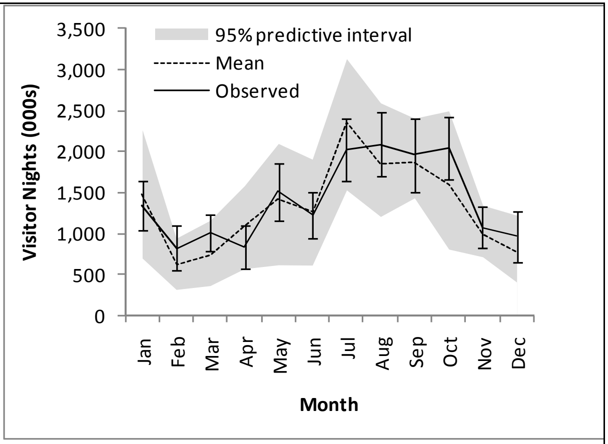


Figure 7: Modelled versus observed data, Northern (Qld)

A similar level and pattern of fit is observed in all but two of the 66 regions for which the model was run. The exceptions were High Country, a region which includes a number of alpine ski resorts (Figure 8) and the Gold Coast, a tourist and urban centre in Queensland (Qld) (Figure 9). In both regions the model underestimates Visitor Nights for a single month. In High Country, Visitor Nights are under predicted in August, corresponding to the peak in the Australian ski season. This appears to be caused by both an under estimation of visit intensity and duration of stay in this region at this time. Visitor Nights are under estimated for the Gold Coast in September due largely to the underestimation of visit durations at this time. In both cases the distributions reference to classes of tourism regions, miss localised deviations from the model profiles.

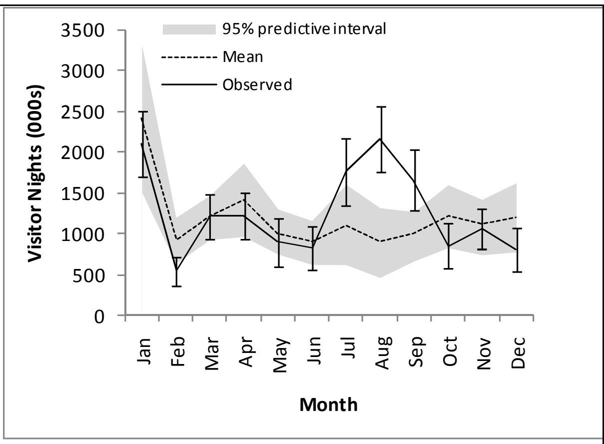


Figure 8: Modelled versus observed values, Visitor Nights, High Country (NSW)

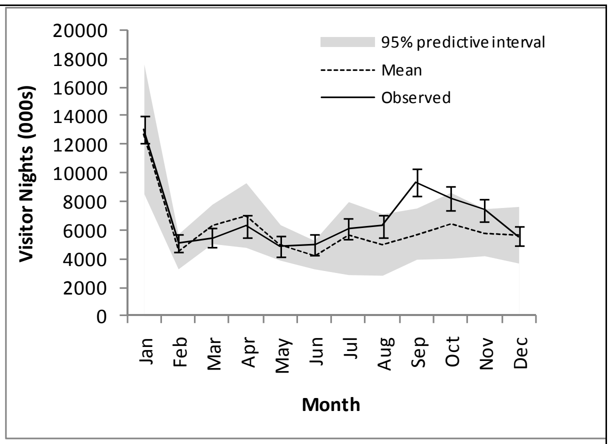


Figure 9: Modelled versus observed values, Visitor Nights, Gold Coast (Qld)

Running the model for an earlier period, 1998-2001, produces similar levels of model fit across the 66 Tourism Regions, with 80.6% of observations for the NVS falling with the 95% predictive interval of modelled data. Again, this fit improves when the sampling variability in the observations is accounted for.

A key impetus in the development of the simulation model for the estimation of visitor populations was to generate estimates at a spatial resolution appropriate for the planning and provision of services. In Australia, this means one of the 667 Local Government Areas (2006) that provide basic services (e.g. waste collection) to both resident and visitor populations. Simulations were run for a selection of 13 Local Government Areas, with the model input parameters specified based on the location of Local Government relative to zonations shown in Figure 2 and 4.

Model estimates are evaluated against hybrid data NVS and Small Area Accommodation data, generated using the Iterative Proportional Fitting (IPF), a widely used method in which disaggregated spatial data are generated from aggregate values (Wong 1992). The mathematical procedure involves the adjustment a 2-dimensional matrix so that the row and column totals equal a pre-defined value. The procedure is defined mathematically as:

$$p_{ij(k+1)} = \frac{p_{ij}(k)}{\sum_{j} {}_{j}p_{ij}(k)} \times Q_i$$

[Equation 6]

$$p_{ij(k+2)} = \frac{p_{ij}(k)}{\sum_{i} p_{ij}(k)} \times Q_j$$

[Equation 7]

Where $p_{ij}(k)$ is the matrix element in row I, and column j for iteration k and Q_i and Q_j are the pre-defined row and column totals respectively. Equations 6 and 7 are repeated until:

$$\sum_j p_{ij}(k) = Q_i \text{ and } \sum_i p_{ij}(k) = Qi$$

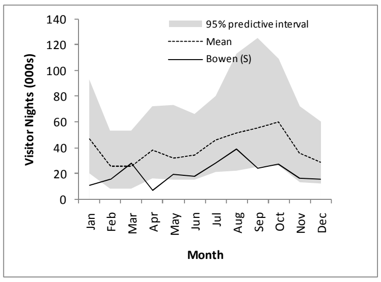
Source: after Norman (1999) and Wong (1992)

Estimates of Visitor Nights for thirteen Local Government Areas for which the model was run are shown in Table 5. The model fit for the thirteen Local Goverenment Areas is similar to what was seen in intial runs of the model for tourism regions with 87.9% of monthly values falling wihin the 95% predictive interval generated from the modelled data. The lack of confidence intervals for the interative proportionally fitted data does not allow futher quantification of investigation of the overall levels of model fit. However insights intomodel performnce can be gained through inspection of the modelled profiles.

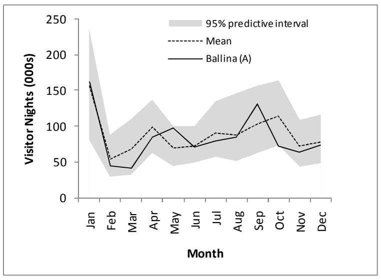
A consequence of shifting to smaller spatial zones, is the appearance of local peaks in Visitor Nights that were smoothed when the model was evaluated at the Tourism Region scale. The profile of Visitor Nights in the Snowy Montains (S), an alpine ski resort, is a good example with peaks over the winter ski season underestimated in the model profile. Another consequence of the shift in spatial scales is the tendency towards over estimation in some regions such as Bowen (S) and Johnstone (S). This appears to be the result of an overestimation of visit duration in these local areas, which are shorter on average than visits to the region as a whole.

Table 5: Model results versus observed data, Visitor Nights, selected Local Government Areas, 2006

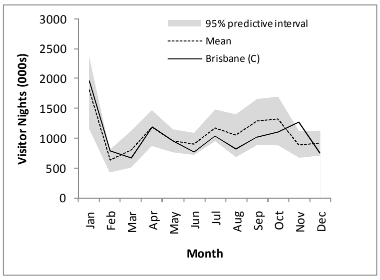
Whitsundays – Bowen(S)



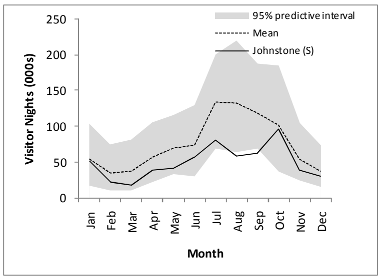
Northern Rivers – Ballina (C)



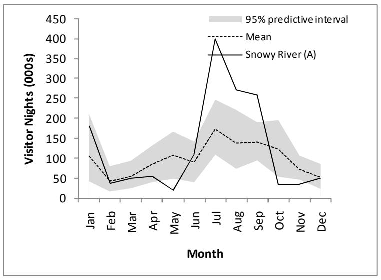
Brisbane – Brisbane (C)



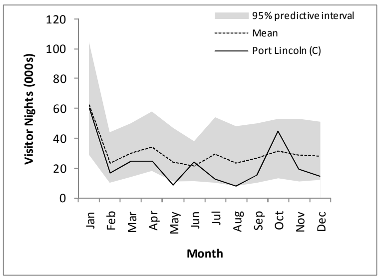
TNQ – Johnstone(S)



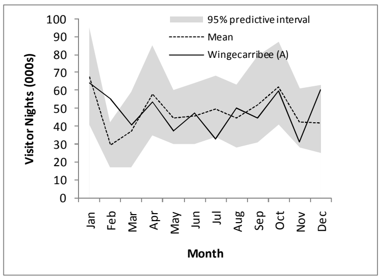
Snowy Mountains - Snowy River (A)



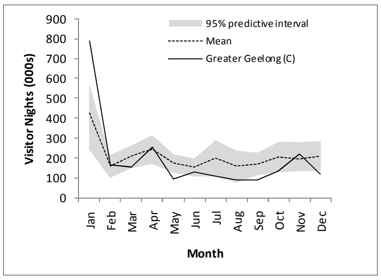
Eyre Peninsula- Port Lincoln (C)



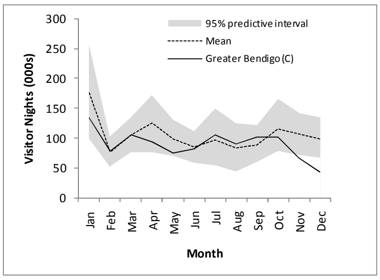
Capital Country-Wingecarribee (A)



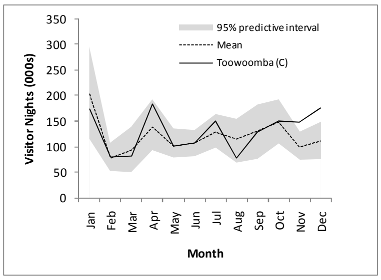
Geelong – Greater Geelong (C)



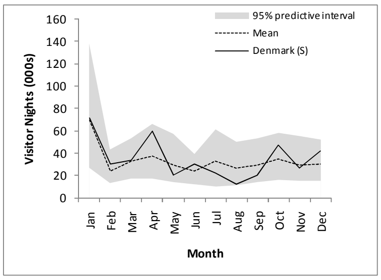
Bendigo-Loddon – Bendigo



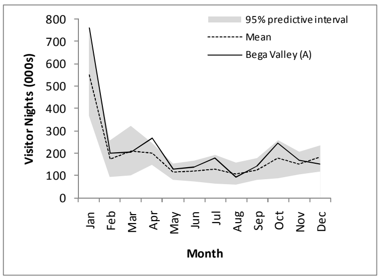
Darling Downs - Toowoomba



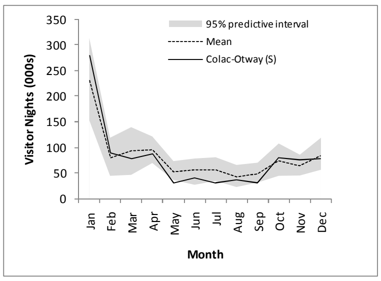
Australia's South West – Denmark



South Coast – Bega Valley (A)



Western – Colac - Otway



Despite these discrepancies between the observed and modelled profiles, model performance at the Local Government Areas level is promising. Further investigation of model performance across a wider array of regions is required.

6. Conclusion

The central thesis of this paper was that variation in temporary visitor populations to Australian destination regions can be estimated by exploiting the spatial and temporal regularities in the Australian temporary population system. This approach stands in contrast to past attempts to estimates temporary populations via ad hoc surveys and symptomatic data that have met with limited success. Model results, for both tourism regions, and Local Government Areas, provide preliminary support for a simulation approach to estimating visitor populations to Australian Regions.

Whilst not to overplay the significance of this research, parallels can be drawn between this initial attempt to model temporary populations via the parameterization of seasonality and duration of visits to region, and the development of fertility, mortality and migration schedules, now widely used in the estimation of resident populations. A key difference is that the regularities exploited here relate to the characteristics of moves, rather than movers. Whilst it is by no means certain that similar regularities exist in other national temporary mobility systems, it is certainly an avenue worth exploring in the pursuit of a broad based methodology for estimating non-resident populations

References

ABS (1996). Review of Demography Statistics. Canberra, Australian Bureau of Statistics.

ABS. (2009). "8635.0.55.002 - Tourist Accommodation, Small Area Data, Australia, Mar 2009 " Retrieved April 2009, 2009, from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/8635.0.55.002Explanatory%20Notes1Mar%202009?OpenDocument.

Australian Bureau of Statistics. (2005). "Tourist Accommodation, Small Area Data, Australia, Jun 2005." Retrieved 06/02/2006, 2006, from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/8635.0.55.002Main+Features1Jun%202005?OpenDocument.

BarOn, R. (1975). Seasonality in Tourism: A Guide to the Analysis of Seasonality and Trends for Policy Making. London, The Economist Intelligence Unit Ltd.

Bell, M. (2004). Measuring temporary mobility: dimensions and issues. CAUTHE, Brisbane.
Bell, M., M. Blake, et al. (2002). "Cross-national comparison of internal migration: issues and measures." Journal of the Royal Statistical Society **165**(3): 435-464.
Bell, M. and G. Ward (1998). "Patterns of temporary mobility in Australia: evidence from the 1991 Census." Australian Geographical Studies **36**(1): 58-81.
Bell, M. and G. Ward (2000). "Comparing temporary mobility with permanent migration." Tourism Geographies **2**(1): 87-107.
Bell, M. and T. Wilson (2004). "Comparative empirical evaluations of internal migration models in subnational population projections." Journal of Population Research **21**(2): 127-160.
Brown, D. and M. Bell (2005). Measuring temporary mobility: dimensions and issues. International Union for the Scientific Study of Population, Tours, France.
Castro, L. J. and A. Rogers (1983). "What the age composition of migrants can tell us." Population Bulletin of the United Nations **15**: 63-79.
Chapman, M. and R. M. Prothero (1983). "Themes on circulation in the third world." International Migration Review **17**(4): 597-631.
de Gans, H. (1994). Some reflections on the concept of migration. Population Dynamics in Europe. P. Hooimeijer, B. van der Knapp, J. van Weesep and R. I. Woods. Utrecht, Royal Netherlands Geographical Society.
Hall, C. M. (2008). Of time and space and other things: laws of tourism and the geographies of temporary mobilities. Tourism and Mobilities: Local=Global Connections. P. M. Burns and M. Novelli. Wallingford, CAB International.
Hanson, J. and M. Bell (2007). "Harvest trails in Australia: patterns of seasonal migration in the fruit and vegetable industry." Journal of Rural Studies **23**: 101-117.
Hogan, T. D. and D. N. Steinnes (1993). "Elderly migration to the sunbelt: seasonal versus permanent." The Journal of Applied Gerontology **12**(2): 246-260.
Hooimeijer, P. and B. van der Knaap (1994). From flows of people to networks of behaviour. Population Dynamics in Europe. P. Hooimeijer, B. van der Knapp, J. van Weesep and R. I. Woods. Utrecht, Royal Netherlands Geographical Society.
Houghton, D. (1993). "Long distance commuting; a new approach to mining in Western Australia." Geographical Journal **159**(3): 281-290.
Judson, D. H. and C. Popoff (2004). Appendix C: Selected General Methods. The Methods and Materials of Demography. J. Siegal and D. A. Swanson. San Diego, Elsevier Axcademic Press.
Koenig-Lewis, N. and E. Bischoff (2005). "Seasonality research: the state of the art." International Journal of Tourism Research **7**: 201-219.
Lee, S. (1999). An Investigation to Assess the Feasibility of Producing Service Population Estimates for Selected LGAs. Canberra, Australian Bureau of Statistics.
Lundtorpe, S. (2001). Measuring tourism seasonality. Seasonality in Tourism. T. Baum and S. Lundtorpe. London, Pergamon.
Mings, R. (1997). "Tracking snowbirds in Australia: winter sun seekers in far North Queensland." Australian Geographical Studies **35**(2): 168-182.
Mooney, C. Z. (1997). Monte Carlo Simulation. Thousand Oaks, California, Sage Publications.
Rees, P. (1997). "Problems and solutions in forecasting geographical populations." Journal of the Australian Population Association **14**(2): 145-166.
Rogers, A., R. Raquillet, et al. (1978). "Model migration schedules and their applications." Environment and Planning A **10**: 475-502.

Rose, L. S. and H. L. Kingma (1989). "Seasonal migration of retired persons: estimating its extent and implications for the state of Florida." Journal of Economic and Social Measurement **15**: 91-104.

Roseman, C. C. (1992). Cyclical and polygonal migration in a Western context. Community, Society and Migration. P. Jobes, W. Stinner and J. Wardwell**:** 33-45.

Smith, S. K. (1989). "Toward a methodology for estimating temporary residents." Journal of the American Statistical Association **84**(406): 430-436.

TRA (2009). Travel by Australians, March 2009. Quarterly Results of the National Visitor Survey. Canberra, Tourism Research Australia.

van Imhoff, E., N. van der Gaag, et al. (1997). "The selection of internal migration models for European regions." International Journal of Population Geography **3**: 137-159.

Wilson, T. and M. Bell (2004). "Comparative empirical evaluations of internal migration models in subnational population projections." Journal of Population Research **21**(2): 127-160.

Wong, D. W. S. (1992). "The Reliability of Using the Iterative Proportional Fitting Procedure." The Professional Geographer **44**(3): 340-348.